User Manual

for

ArthropodEST pipeline System

Version 1.0

Submitted in partial fulfillment of the requirements of the degree of Master of Software Engineering

Prepared by

Luis Fernando Carranco M.
CIS 895 – MSE Project
Kansas State University
Spring 2010

Table of Contents

1 Introduction 6

2 Installation and Configuration 6

2.1 Bioinformatics Server 6

2.1.1 Required Software 6

2.1.2 Database 7

2.1.3 Engine coordinator 7

2.1.4 Web Interface 10

2.2 Beocat Cluster Server 11

2.2.1 Event Log component 12

2.2.2 Bioinformatics Software 12

2.2.3 Pipeline scripts 13

3 ArthropodEST Pipeline Usage 14

3.1 User Interface 14

3.1.1 Submit a Request 15

3.1.2 Monitor Request 16

3.1.3 Cancel Request 17

3.1.4 View Request Results 18

3.2 Administrator Interface 19

3.2.1 Login 20

3.2.2 Logout 20

3.2.3 Change Password 20

3.2.4 Resources 21

3.2.5 Reports 22

3.2.6 Analysis Request Detail 24

3.2.7 Cancel Requests 26
3.2.8 View Request Results 27
3.2.9 Start and Stop Engine coordinator 27

Table of Figures

Figure 1 Request submission Web Page of ArthropodEST pipeline system 16

Figure 2 Monitor a request Web Page of ArthropodEST pipeline system 17

Figure 3 Cancel a request Web Page of ArthropodEST pipeline system 18

Figure 4 Results Web Page of ArthropodEST pipeline system 19

Figure 5 Login Web Page. Administration Interface 20

Figure 6 Change Password Web Page. Administration Interface 21

Figure 7 Resource Web Page. Administration Interface 22

Figure 8 Reports Web Page. Administration Interface 24

Figure 9 Request Detail Web Page. Administration Interface 25

Figure 10 Cancel Request. Administration Interface 26

Table of Tables

- Table 1 Configuration file engine-est.cfg 10
- Table 2 Configuration file engine-logging.cfg 10
- Table 3 Configuration file log-est.cfg 12

1 Introduction

This document provides a user manual for ArthorpodEST Pipeline System. It is divided in two sections: Installation and Configuration, and System Usage. The first section describes the components to be installed and the possible configurations supported. The next section, Usage, describes how to interact with the system from the user and administrator perspective.

The intention of this manual is to guide experienced researchers and technicians familiar with expressed sequence tag (EST) analysis and correspondent Bioinformatics Software. It is expected that users have knowledge of the EST pipeline analysis process. This manually, simply explains the operation of ArhtorpodEST Pipeline software components and does not go into details about the bioinformatics processing activities.

2 Installation and Configuration

ArthropodEST Pipeline is a distributed system that needs different components to be installed over two main servers: Bioinformatics Center (BC) and Beocat Cluster Server (refer to Architecture Design document, section 2.1 for more information on physical distribution)

2.1 Bioinformatics Server

The following components of ArthropodEST pipeline system will be installed on this server:

- Database. The centric information repository of the Pipeline system.
- User Interface. Web pages to interact with users.
- Engine component. Coordinator to manage and processes requests.

2.1.1 Required Software

- GNU Linux Ubuntu Server 8.04 or superior

- MySQL 5.0 or superior
- Python 2.6 including python-mysqldb, pycryto, paramiko libraries
- Perl 5.0 including CGI, File::Temp, DBI, DBD::mysql libraries
- Php 5.0 including php5-mysql library
- Apache 2.2 including mod_perl, mod_php add-ins

2.1.2 Database

The main system database use MySQL as Relational Database Engine. The rest of components will need to set up a connection configuration string according where this database will be mounted. It is important to notice that it might be installed on a separated machine as a dedicated server. Below are the steps to set up a database. MySQL root user must perform the installation.

- Create a database for ArthropodEST Pipeline system

```
CREATE DATABASE arthropodest;
```

- Create a user with permissions to access the database from BC server and Beocat. Their IP addresses are needed for set up this

```
CREATE USER 'arthropodest'@'bcserverip' IDENTIFIED BY 'athropodestpass';
CREATE USER 'arthropodest'@'beocatip' IDENTIFIED BY 'athropodestpass';
GRANT SELECT,INSERT,UPDATE,DELETE,TRIGGER USER ON arthropodest.* TO
'arthropodest'@'bcserverip';
GRANT SELECT,INSERT,UPDATE,DELETE,TRIGGER USER ON arthropodest.* TO
'arthropodest'@'beocat';
```

- Create database structure. Use the schema.sql file provided on the system package.

```
mysql -uroot -ppassroot -hlocalhost -P 3306 dbName < /path/to/schema.sql
```

2.1.3 Engine coordinator

The engine coordinator is a service coded in python and it runs as a Unix daemon. It must be launched as a root because the service later on will automatically drop privileges to run as the apache

user ' www-data'. The following are the steps to set up the engine. Root user must perform the installation.

- Unpack the engine source code in a desired location on server
- Grant 'read' and 'write' permissions only for root user to configurations files: engine-est.cgf and engine-logging.cfg
- Modify configuration files according the parameters specified in tables below. Notice that configuration files follows the python convention grouping parameters entries `name = value` in sections headers `[section]` (http://docs.python.org/library/configparser.html). Do not change parameter or section names on those files, but values according necessities.

Section	Parameter Name	Format	Default / Example	Description
[MySql]	host	STRING	localhost	Database host name (might be IP address)
	port	INT	3306	Database connection port
	user	STRING	dbuser	Database user name
	password	STRING	dbpassword	Database password
	db	STRING	dbname	Database name
[Beocat]	host	STRING	beocat.cis.ksu.edu	Beocat host name (might be IP address)
	user	STRING	bioinfo	Beocat user account where pipeline components resides
	password	STRING		Beocat user password. This argument is optional if a private key is used to authenticate.
	rsa_file	STRING	/home/id_rsa	Local absolute path where private key of Beocat user is located. If specified, the correspondent public key must be included on Beocat user account file ~/.ssh/authorized_keys
	rsa_password	STRING		If the private key provided is encrypted, the password to decrypt it must be defined here.
[SMTP Server]	Server	STRING	localhost	SMTP server host name (might be IP address)
	Port	INT	587	SMTP server port number
	Email	STRING	user@localhost.com	User login name for SMTP server. Must be defined if server requires authentication
	password	STRING		User password for SMTP server. Must be defined if server requires authentication
[Scheduler]	maxConcurrentJobs	INT	20	Number of maximum allowed executions that can run concurrently on Beocat.

	maxCpus	INT	50	Number of maximum allowed cpus that can be used concurrently on Beocat.
	threshold	INT	50	Percentage of simultaneous executions on Beocat that need to be overcome before using the minimum resources to allocate new executions. Otherwise, the maximum resources will be allocated for each execution. Refer 'resource administration ' on usage section.
	maxCpusPerJob	INT	16	Number of maximum cpus that can be used per each execution on Beocat. This must match the maximum number of cpus that one node on Beocat supports. Executions cannot share cpus from different nodes.
[Mail Recipients]	sender	STRING	bioinfo@ksu.edu	Email of the sender of notification emails
	admin	STRING	me@ksu.edu	Administrator email(s)
	cc	STRING		Email(s) that will be carbon copied for all the notifications to users
	bcc	STRING		Email(s) that will be blind carbon copied for all the notifications to users
[General]	fileParameters	STRING	arguments	File name that will content the arguments built by web page request (e.pl)
	wwwUploadBaseDir	STRING	/tmp/www_e	Local folder where apache will upload the input files of requests
	remoteFolder	STRING	arthropodest	Name of remote folder on Beocat where the pipeline components where installed
	localFolder	STRING	/var/www/e	Local folder where the web pages for user requests are located
	jobScript1	STRING	~/ job-e-seqcl-repmask.sh	Path where the script to run job 1 is located on Beocat (sequence cleaner, repmasker, and cap3 executions)
	jobScript2	STRING	~/job-e-signalp.sh	Path where the script to run job 2 is located on Beocat (signal executions)
	jobScript3	STRING	~/job-e-blast2go.sh	Path where the script to run job 3 is located on Beocat (blast2go executions)
	email_ack	STRING	email_ack	File name that will content the email content that will be built by web page request (e.pl)
	urlHost	STRING	http://129.130.115.77/e	Base URL where the web interface can be browsed
	timeMonitor	INT	5	Seconds for the frequency that the engine will process the queue
	totalRuns	STRING	/var/www/e/log/total_runs	Path to the file that will content the number of total requests processed by the system
	queueLength	STRING	/var/www/e/log/queue_length	Path to the file that will content the number of current request being processed

	daemon_stdout	STRING	/tmp/est.out	Path to the file where engine stdout will be redirected (/dev/null by default)
	daemon_stderr	STRING	/tmp/est.err	Path to the file where engine stderr will be redirected (/dev/null by default)
	daemon_pidfile	STRING	/tmp/est.pid	Path to the file that will hold the engine process id. This will be present while engine running.

- **Table 1 Configuration file engine-est.cfg**

Logging configuration file establishes how the engine coordinator will log the events generated during the processing of a request. For more information how to add more handlers and change format of output refer python documentation (http://docs.python.org/library/logging.html). There are two loggers used by the application "Engine" and "Paramiko" Log Handlers. The recommended parameters that might be changed are listed on the table below.

Section	Parameter Name	Format	Default / Example	Description
[logger_engine-est]	level	STRING	INFO	Level of messages to be logged: DEBUG, INFO, ERROR
	handlers	STRING	fileHandler	Handlers used to manage engine log messages
[logger_paramiko]	level	STRING	ERROR	Level of messages to be logged: DEBUG, INFO, ERROR
	handlers	STRING	fileHandler_paramiko	Handlers used to manage paramiko library log messages
[handler_fileHandler]	args	STRING	('/tmp/est.log','a')	Engine file handler. Specify the log file path for the main application messages
[handler_fileHandler_paramiko]	args	STRING	('/tmp/paramiko.log','a')	Paramiko Library file handler. Define log file path for paramiko library's messages

- **Table 2 Configuration file engine-logging.cfg**

2.1.4 Web Interface

The web interface will content the web pages to interact with user and administrator. To install the web interface the following steps needs to be done:

- Unpack the web pages source code in a folder with permissions access for apache user
- Configure the folder as a virtual directory using apache configuration manual.
- Edit admin/config.php file and configure the following lines of code to be able to connect ArthropodEST MySQL database:

```
$database = 'arthropodest';
$user = 'arthropodest';
$pass = 'arthropodestpass';
$host = 'dbhostname';
$port = 3306;
```

- Edit e.pl file and configure the following lines of code in its header with appropriate values:

```
#DATABASE Section
my $db="arthropodest";
my $host="localhost";
my $port=3306;
my $dbuser="arthropodest";
my $dbpass="arthropodestpass";
#GENERAL Section
my $URL = "http://129.130.115.77/e2-beocat";
my $REMOTE_EST_FOLDER = "~/arthropodest";
my $LOCAL_EST_UPLOAD_DIR = "/tmp/e2-beocat";
#ANTIVIRUS Section
my $CLAMSCAN = "/usr/bin/clamscan";
my $CLAMSCAN_OPT="--quiet";
```

- Edit e.cgi file and configure the following lines of code in its header with appropriate values:

```
TOTAL_RUNS_FILE=/var/www/e2-beocat/log/total_runs
QUEUE_LENGTH_FILE=/var/www/e2-beocat/log/queue_length
```

Notice that most of the variable values are the same as the main engine configuration file. Those variable values should match in order to execute the components.

2.2 Beocat Cluster Server

The following components of ArthropodEST pipeline system will be installed on this server:

- Event log component to keep track of errors and status of analysis execution.
- Pipeline scripts to perform the analysis requests.
- Bioinformatics software.

2.2.1 Event Log component

This component allows the interaction between pipeline scripts on Beocat and main engine coordinator on BC Server. It is a script developed in python. The following are the steps to set up the Event log script. Beocat account user must perform the installation.

- Unpack the Event log source code in a sub folder of Beocat account user.
- Grant 'read' and 'write' permissions to configuration files to only Beocat account user.
- Make the modifications on the configuration file `log-est.cfg` to communicate with the main System Database on BC Server according to table below. Notice that the configuration file is a small version of the configuration file of the main engine application.

Section	Parameter Name	Format	Default / Example	Description
[MySql]	host	STRING	localhost	Database host name (might be IP address)
	port	INT	3306	Database connection port
	user	STRING	dbuser	Database user name
	password	STRING	dbpassword	Database password
	db	STRING	dbname	Database name

- **Table 3 Configuration file log-est.cfg**

2.2.2 Bioinformatics Software

Bioinformatics programs must be installed and able to run on Beocat user account. The necessary packages which are launched from pipeline scripts and need to be installed are listed below:

- TGICL software.
- SeqClean.
- NCBI BLAST suite.

- Vectors databases.
- RepeatMasker software, including Tandem Repeats Finder, Repeat Database, cross_match.
- CAP3
- Blast2Go
- Signal
- EMBOSS

2.2.3 Pipeline scripts

The pipeline scripts are the main code to perform the EST analysis. They will execute the Bioinformatics Software and Event Log component previously installed. The following are the steps to set up the Pipeline scripts. Beocat account user must perform the installation.

- Unpack the pipeline scripts source code in a sub folder of Beocat account user.
- Grant 'read' and 'write' permissions only Beocat account user.
- Make the modifications on the configuration file `ArthropodEST_conf.sh` to determine where the bioinformatics software and event log component are installed. Modify the following lines of code as necessary:

```
WWW_HOST_URL='http://129.130.115.77/e2-beocat'
LOG=~/arthropodest/src/log-est.py #Event log component
WWW_EST_LOG_DIR=~/arthropodest/log # full PATH of log directory
#Location Bioinformatics Software
SCRIPTS=~/arthropodest/bin
SEQSTATS_WORKDIR=~/arthropodest/tmp
NCBI_UNIVEC=~/bioinfo_software/vectors/ncbi_univec/UniVec_Core
EMBL_EMVEC=~/bioinfo_software/vectors/embl_emvec/emvec.dat.fsa
FORMATDB=~/bioinfo_software/ncbi-blast/bin/formatdb
SEQCL=~/bioinfo_software/tgi/seqclean/seqclean
```

```
CLN2QUAL=~/bioinfo_software/tgi/seqclean/cln2qual
CDBYANK=~/bioinfo_software/tgi/seqclean/bin/cdbyank
REPMASK=~/bioinfo_software/RepeatMasker/RepeatMasker
TGICL=~/bioinfo_software/tgi/tgicl_linux/tgicl
CAP3=~/bioinfo_software/cap3/cap3
BLASTX='/homes/bioinfo/bioinfo_software/ncbi-blast/netblast/bin/blastcl3 -p
blastx -d nr'
BLASTX_PARAMS='-e 1e-04 -m 7 -b 20 -v 20'
JAVA=/usr/bin/java
JAVA_OPTS='-client -Xms64m -Xmx512m'
BLAST2GO="${JAVA} ${JAVA_OPTS} -jar
/homes/bioinfo/bioinfo_software/blast2go/blast2go.jar -prop
/homes/bioinfo/bioinfo_software/blast2go/b2gPipe.properties"
BLAST2GO_PARAMS='-a -d'
AWK=/usr/bin/awk
PYTHON=/usr/bin/python
# SET PATH FOR RUNNING ON BEOCAT NODES
ArthropodEST_PATH=~/bioinfo_software:~/bioinfo_software/tgi:~/bioinfo_software/
tgi/tgicl_linux:~/bioinfo_software/tgi/tgicl_linux/bin:~/bioinfo_software/tgi/s
eqclean:~/bioinfo_software/tgi/seqclean/bin:~/bioinfo_software/ncbi-
blast/bin:~/bioinfo_software/ncbi-blast/netblast/bin:~/bioinfo_software/wu-
blast:~/bioinfo_software/RepeatMasker:~/bioinfo_software/cap3:~/bioinfo_softwar
e/pcap.rep:~/bioinfo_software/phred_phrap_consed:~/bioinfo_software/signalp:~/b
ioinfo_software/blast2go:~/bioinfo_software/GFF3-
Validator:~/bioinfo_software/EMBOSS/bin:~/arthropodest/bin:~/bin
export PATH=${PATH}:$ArthropodEST_PATH
```

3 ArthropodEST Pipeline Usage

3.1 User Interface

The purpose of this section is to provide a reference how a user must interact with the ArthropodEST Pipeline system to perform analysis requests.

3.1.1 Submit a Request

User should open the request web page inside a browser using the appropriate URL. For example: http://arthropodest.bioinformatics.ksu.edu/e-beocat. The following screen will be displayed:

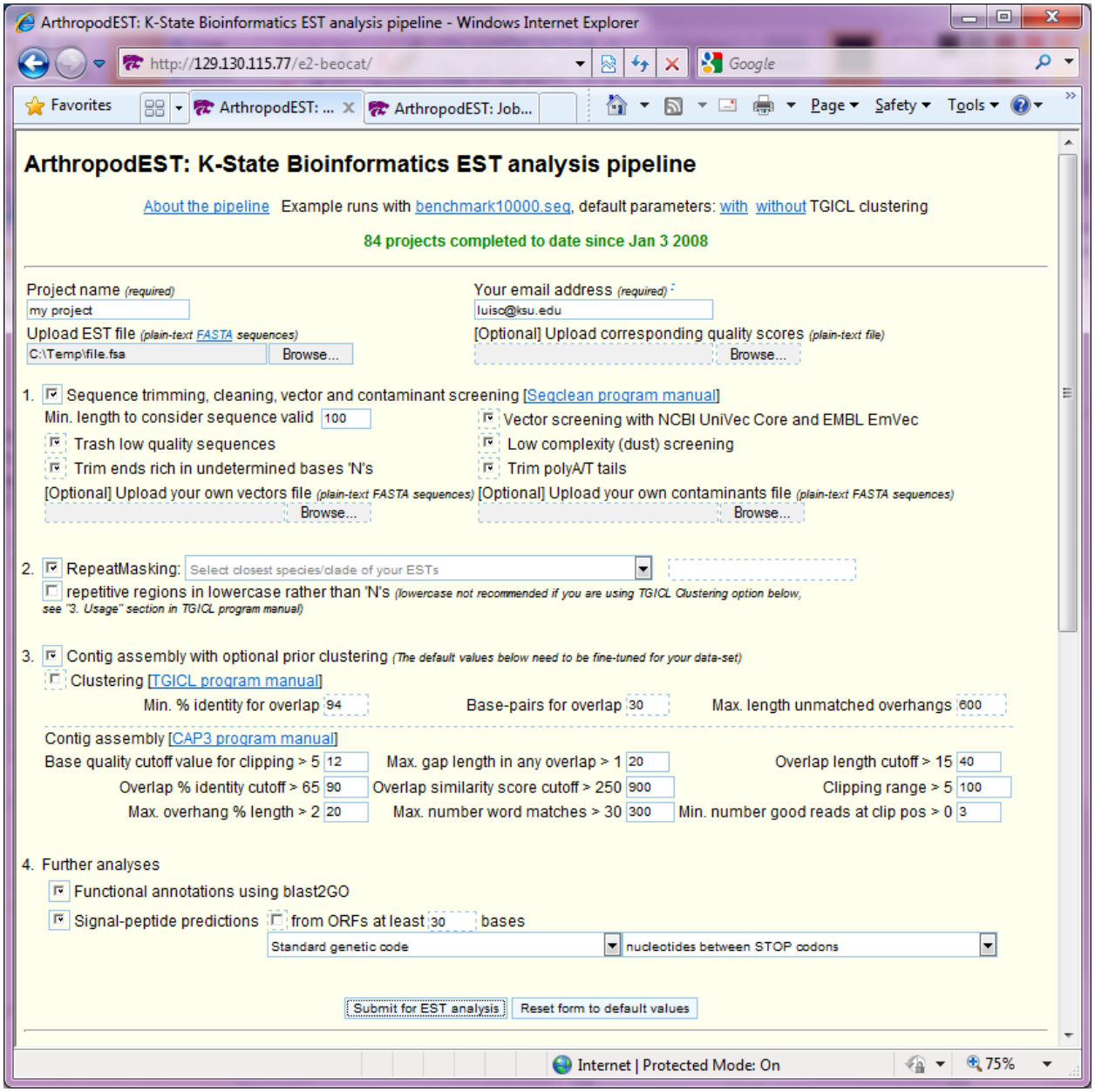
ArthropodEST: K-State Bioinformatics EST analysis pipeline - Windows Internet Explorer
http://129.130.115.77/e2-beocat/
Google
Favorites
ArthropodEST: ...
ArthropodEST: Job...
Page
Safety
Tools
ArthropodEST: K-State Bioinformatics EST analysis pipeline
About the pipeline Example runs with benchmark10000.seq, default parameters: with without TGICL clustering
84 projects completed to date since Jan 3 2008
Project name (required)
my project
Your email address (required)
luisc@ksu.edu
Upload EST file (plain-text FASTA sequences)
C:\Temp\file.fsa
Browse...
[Optional] Upload corresponding quality scores (plain-text file)
Browse...
1. Sequence trimming, cleaning, vector and contaminant screening [Seqclean program manual]
Min. length to consider sequence valid 100
Trash low quality sequences
Trim ends rich in undetermined bases 'N's
Vector screening with NCBI UniVec Core and EMBL EmVec
Low complexity (dust) screening
Trim polyA/T tails
[Optional] Upload your own vectors file (plain-text FASTA sequences)
Browse...
[Optional] Upload your own contaminants file (plain-text FASTA sequences)
Browse...
2. RepeatMasking: Select closest species/clade of your ESTs
repetitive regions in lowercase rather than 'N's (lowercase not recommended if you are using TGICL Clustering option below, see "3. Usage" section in TGICL program manual)
3. Contig assembly with optional prior clustering (The default values below need to be fine-tuned for your data-set)
Clustering [TGICL program manual]
Min. % identity for overlap 94
Base-pairs for overlap 30
Max. length unmatched overhangs 600
Contig assembly [CAP3 program manual]
Base quality cutoff value for clipping > 5 12
Max. gap length in any overlap > 1 20
Overlap length cutoff > 15 40
Overlap % identity cutoff > 65 90
Overlap similarity score cutoff > 250 900
Clipping range > 5 100
Max. overhang % length > 2 20
Max. number word matches > 30 300
Min. number good reads at clip pos > 0 3
4. Further analyses
Functional annotations using blast2GO
Signal-peptide predictions from ORFs at least 30 bases
Standard genetic code
nucleotides between STOP codons
Submit for EST analysis
Reset form to default values
Internet | Protected Mode: On
75%

Figure 1 Request submission Web Page of ArthropodEST pipeline system

First, users must enter the project name, an email address to receive notification about results, and an input EST file. After that, they must select at least one of the possible analysis operations shown in the web page. Every operation is linked to one correspondent Bioinformatics Software. They may choose appropriate parameters for each option according their necessities. The reference about the programs to be launched for EST analysis is described at the bottom of the web page. Users are supposed to understand the analysis operations for their EST input files. Finally, users press 'Submit for EST analysis' button and wait for being notified by email when results are ready.

After a request is placed, if Beocat has available resources and currently there are not running the maximum number of allowed executions, the engine will submit correspondent jobs to the cluster and a first email notification is sent to user to determine the analysis has begun. Notice that even the process was already sent to Beocat, the execution can be delayed by Beocat scheduler because it might be very busy with no nodes available to start the execution.

3.1.2 Monitor Request

The first notification email will include a URL to a web page so that users may monitor or cancel the analysis request. The URL will include a unique identifier correspondent to particular request. For example: http://129.130.115.77/e2-beocat/myjob.php?project=my_project.Rsj3YpmOpRH5

Users might monitor their requests opening that web page on a browser. The status of the selected analysis operations will be shown in a detailed table. Figure 2 shows a screen shot with a request that was recently placed and waiting to be executed. The 'Refresh Information' button allows user reloads the page to show the latest information regarding the status of the analysis request.

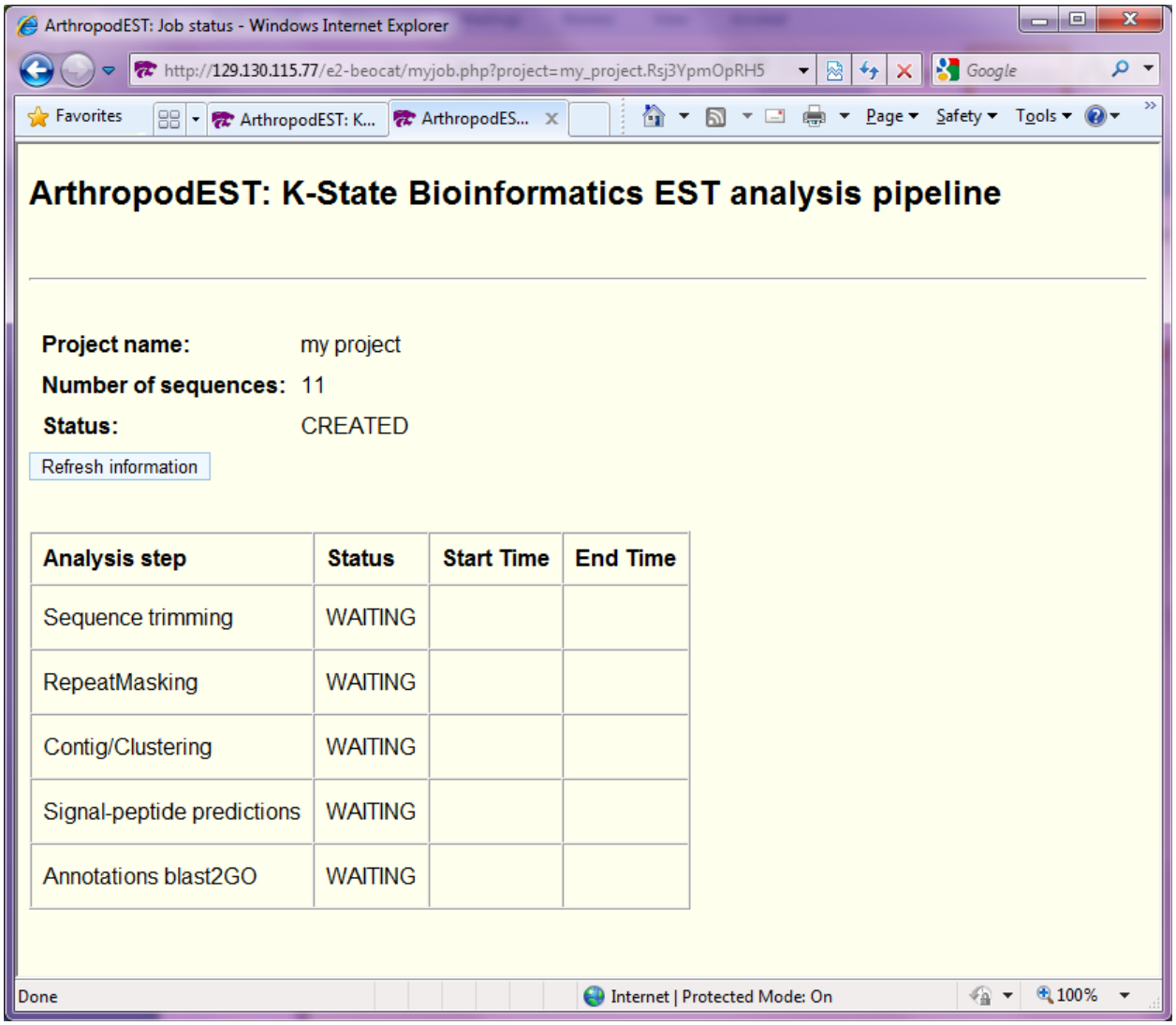

Figure 2 Monitor a request Web Page of ArthropodEST pipeline system

3.1.3 Cancel Request

On the status page, if a request is processing, it will appear a 'Cancel Request' button that allow users to cancel the request execution. After pressing the button, a message will appear so that users confirm that they actually want to cancel the request. After cancelling, a notification will be sent to user to indicate that the request was successfully cancelled. The following screen shot shows an executing request that might be cancelled.

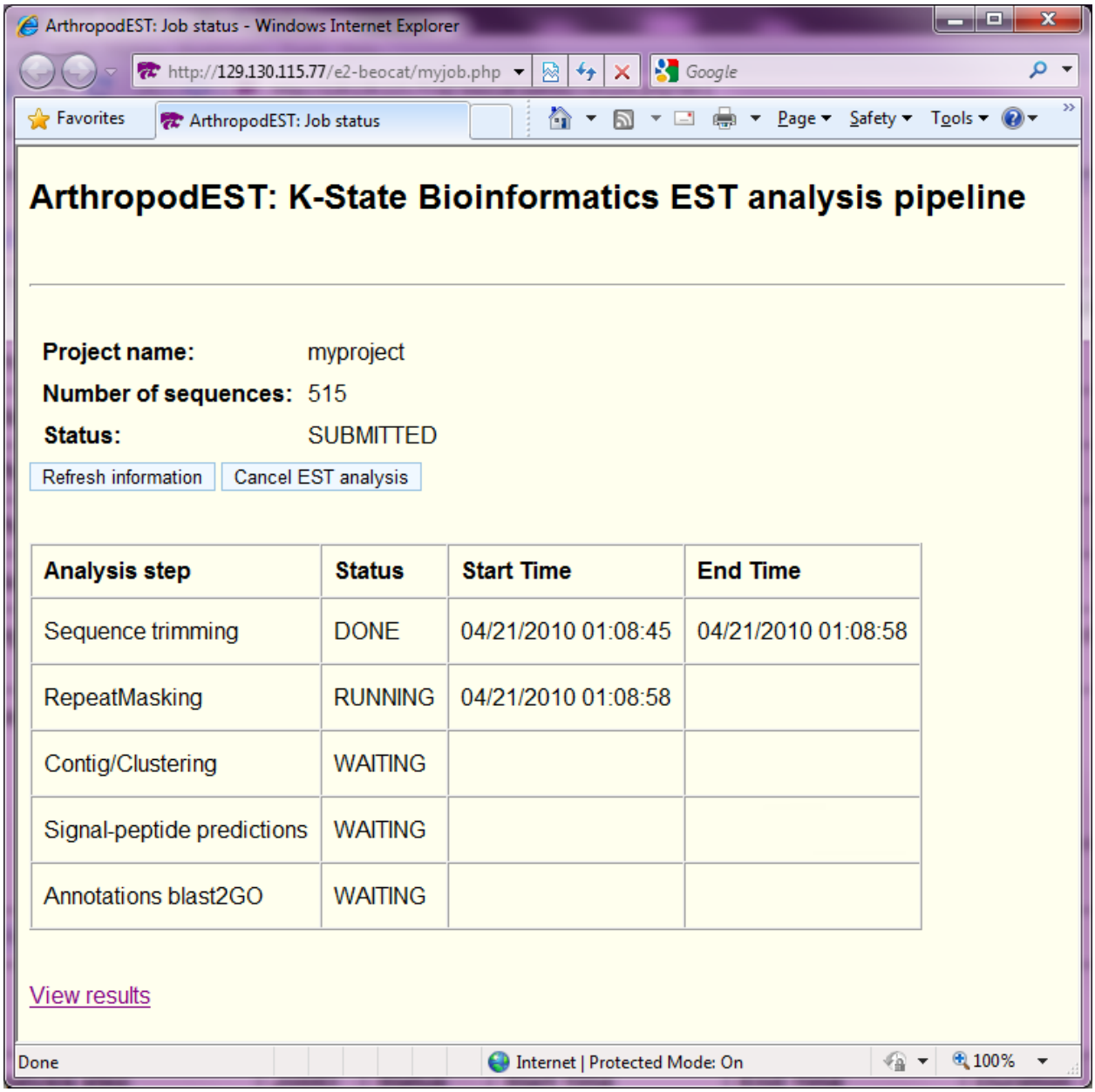

Figure 3 Cancel a request Web Page of ArthropodEST pipeline system

3.1.4 View Request Results

When results are ready, the system will send a notification email with a link to a web page where the user can download the results of the completed analysis. It is important to notice that depending on the options that the user selected to perform the analysis, the system may send two emails notifying the results. This is because the system will execute the analysis in two phases and after each of them it will

retrieve the results from Beocat and send the notification to the User. The options for first phase are the Bioinformatics Programs: Sequence cleaner, repmasker, and CAP3. On second phase Blast2go and signal Peptides predictions are executed.

Finally, on monitor status Web Page, if results are ready to download, a link in the bottom of the page will allow users to open the same Results Web Page. The following is an example of result page.

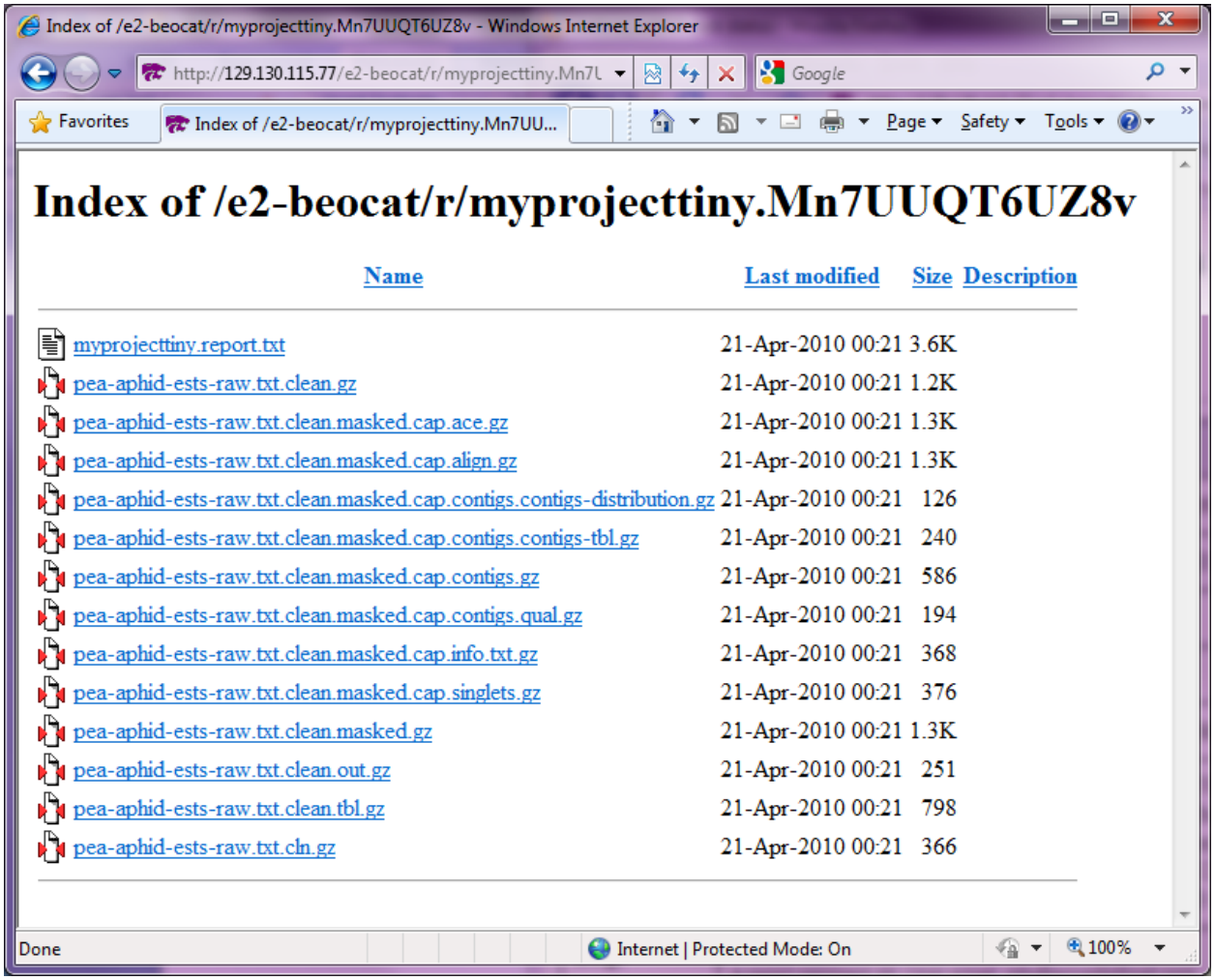

Figure 4 Results Web Page of ArthropodEST pipeline system

3.2 Administrator Interface

The purpose of this section is to provide a guide to administrator(s) about how they must interact with the ArthropodEST Pipeline system to perform management activities.

3.2.1 Login

The login web page is intended to authenticate the administrator user before using the administrator Web Pages. User must enter the username 'admin' and the correspondent password. The default password is 'admin'.

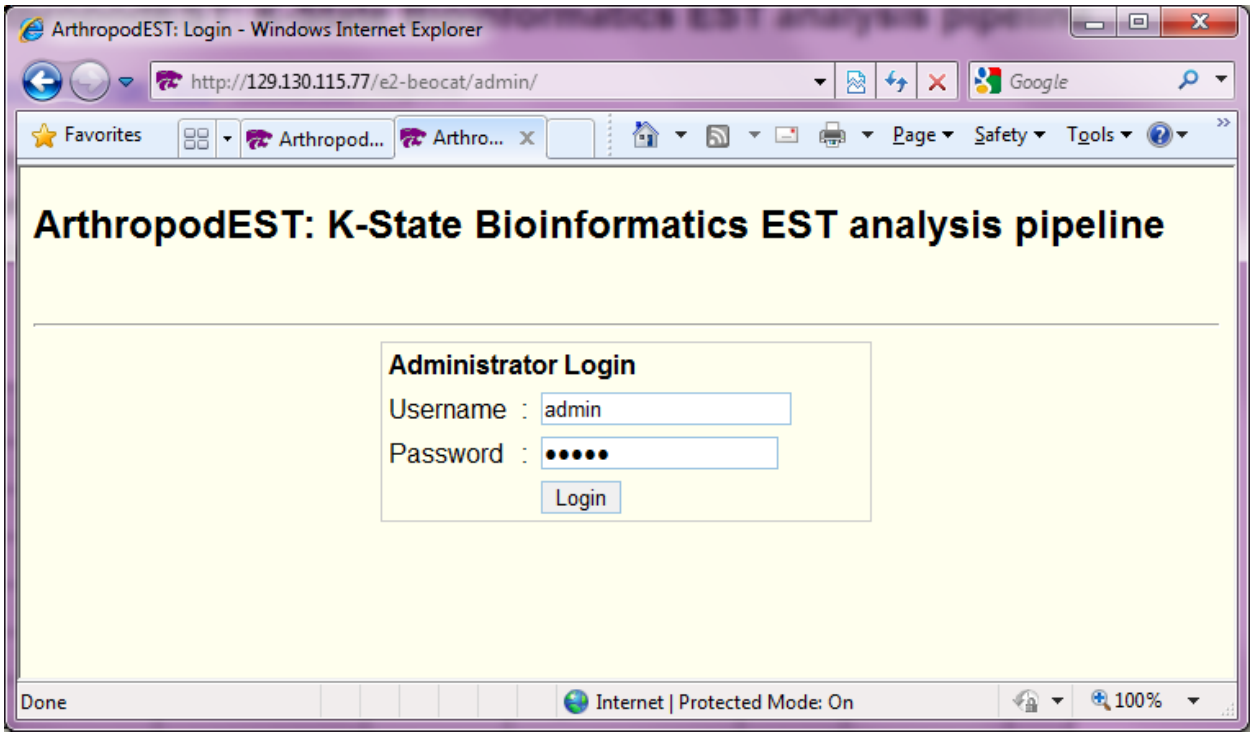

Figure 5 Login Web Page. Administration Interface

3.2.2 Logout

When the administrator was authenticated, he or she can exit from the session using the logout link provided on the administrator interface at the left-top. The administrator might logout any time during the administration activities. If the administrator does not do anything during 15 minutes, the session will automatically expire. The time out configuration can be done on section sessions on `/etc/php.ini` file.

3.2.3 Change Password

This Web Page is intended to change the password for the administrator. He or she needs to enter the old password and the new password twice. The following figure shows the interface for changing password admin.

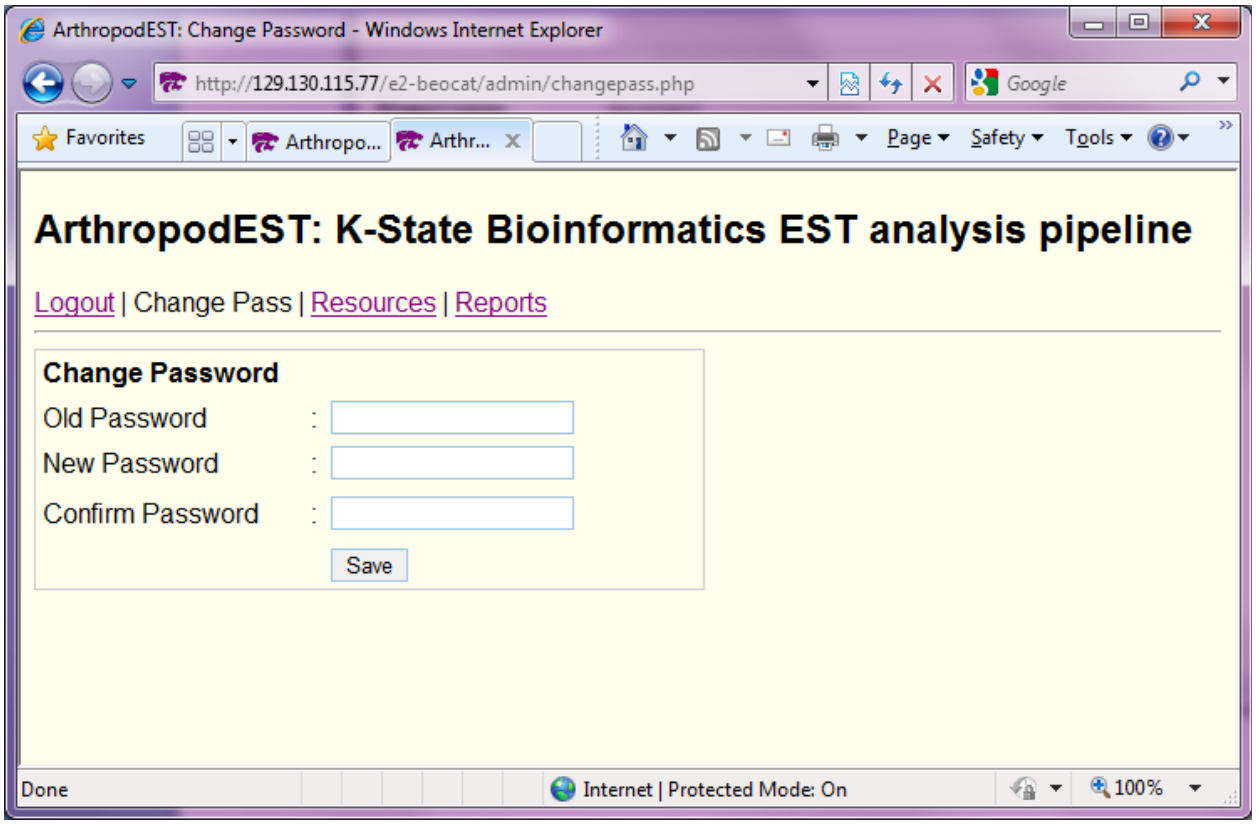

Figure 6 Change Password Web Page. Administration Interface

3.2.4 Resources

This Web Page allows administrator to configure how the engine scheduler will allocate resources to requests. The key factor will be the number of sequences to be analyzed on an input file. The scheduler will scan the resource records from the smallest sequence value to the largest one until it finds the correspondent match for a particular request. For every request there must be 3 records corresponding to the 3 pipeline scripts that can be executed on Beocat. Before allocating the number of CPUs, amount of memory and hours for a particular execution, the engine will count the number of currently executions running and compute the percentage of charge according the maximum executions

allowed. Then, if the current percentage is less or equal than the threshold configured, it will allocate the maximum values. Otherwise, the minimum values will be used to allocate the request. For more information about configuration parameters look at configuration engine file at section 2.3.1.

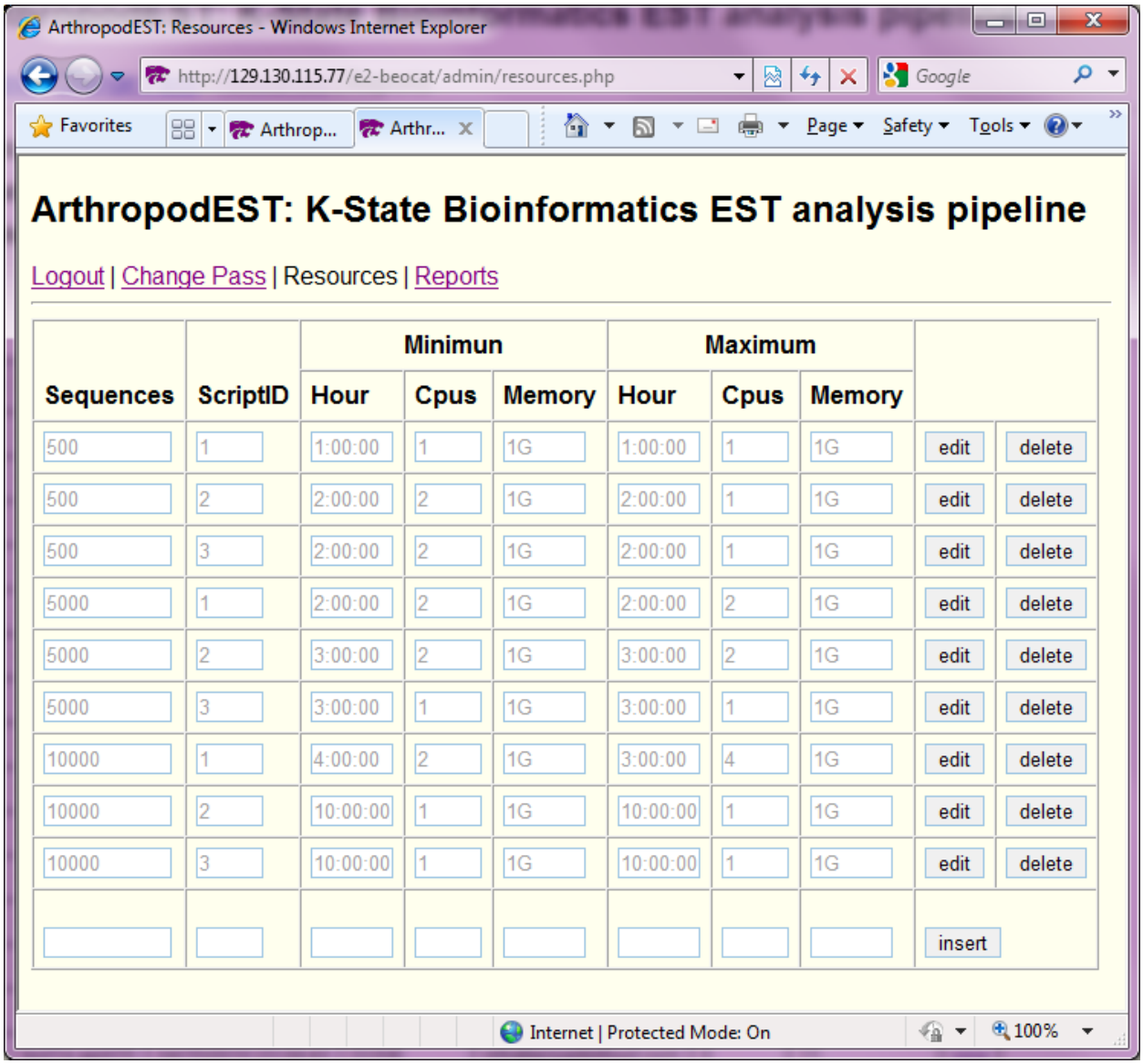
ArthropodEST: Resources - Windows Internet Explorer
http://129.130.115.77/e2-beocat/admin/resources.php
Google
Favorites
Arthrop...
Arthr...
Page
Safety
Tools
ArthropodEST: K-State Bioinformatics EST analysis pipeline
Logout | Change Pass | Resources | Reports
Minimun
Maximum
Sequences
ScriptID
Hour
Cpus
Memory
Hour
Cpus
Memory
500 1 1:00:00 1 1G 1:00:00 1 1G edit delete
500 2 2:00:00 2 1G 2:00:00 1 1G edit delete
500 3 2:00:00 2 1G 2:00:00 1 1G edit delete
5000 1 2:00:00 2 1G 2:00:00 2 1G edit delete
5000 2 3:00:00 2 1G 3:00:00 2 1G edit delete
5000 3 3:00:00 1 1G 3:00:00 1 1G edit delete
10000 1 4:00:00 2 1G 3:00:00 4 1G edit delete
10000 2 10:00:00 1 1G 10:00:00 1 1G edit delete
10000 3 10:00:00 1 1G 10:00:00 1 1G edit delete
insert
Internet | Protected Mode: On
100%

Figure 7 Resource Web Page. Administration Interface

3.2.5 Reports

This administration section is intended to retrieve from MySQL database the analysis requests that have been placed and processed by the ArthropodEST Pipeline System. The header provides different

alternatives to filter the search. The footer provides navigation for the pagination of the list of analysis requests. Figure 8 shows how a report looks like including the table that lists the analysis requests with the main information. To review the complete information of a specific analysis request, the administrator can select the link view at the right of every request. The possible filters to search requests are: Job ID (job identifier assigned by Beocat), project name, user email, status (CREATED, CANCELLED, SUBMITTED, STOPPED, DONE, ERROR), and range of dates when requests were placed.

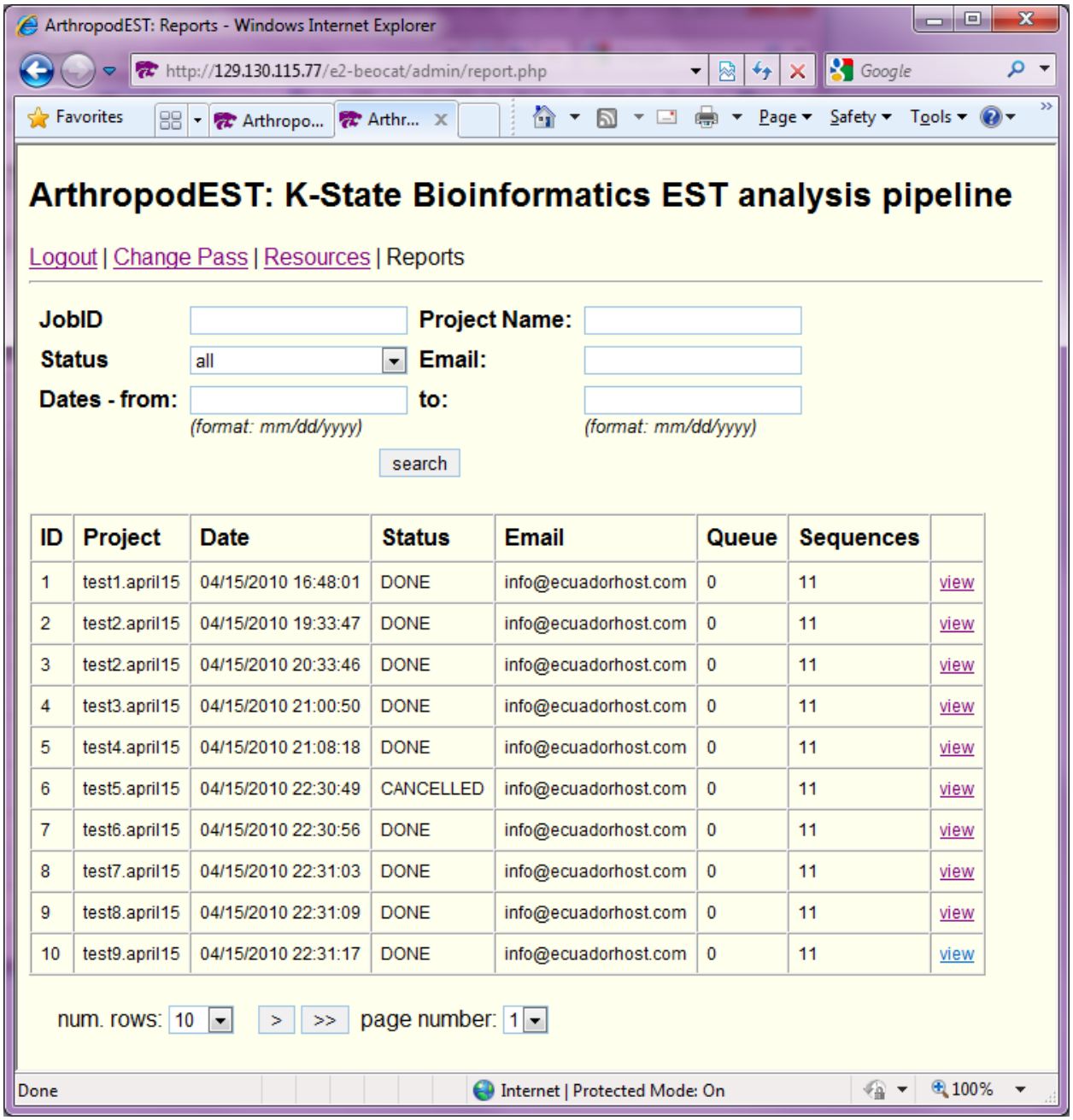
ArthropodEST: Reports - Windows Internet Explorer
http://129.130.115.77/e2-beocat/admin/report.php
Favorites
Arthropo...
Arthr...
Page
Safety
Tools
ArthropodEST: K-State Bioinformatics EST analysis pipeline
Logout | Change Pass | Resources | Reports
JobID
Project Name:
Status
all
Email:
Dates - from:
to:
(format: mm/dd/yyyy)
(format: mm/dd/yyyy)
search
ID Project Date Status Email Queue Sequences
1 test1.april15 04/15/2010 16:48:01 DONE info@ecuadorhost.com 0 11 view
2 test2.april15 04/15/2010 19:33:47 DONE info@ecuadorhost.com 0 11 view
3 test2.april15 04/15/2010 20:33:46 DONE info@ecuadorhost.com 0 11 view
4 test3.april15 04/15/2010 21:00:50 DONE info@ecuadorhost.com 0 11 view
5 test4.april15 04/15/2010 21:08:18 DONE info@ecuadorhost.com 0 11 view
6 test5.april15 04/15/2010 22:30:49 CANCELLED info@ecuadorhost.com 0 11 view
7 test6.april15 04/15/2010 22:30:56 DONE info@ecuadorhost.com 0 11 view
8 test7.april15 04/15/2010 22:31:03 DONE info@ecuadorhost.com 0 11 view
9 test8.april15 04/15/2010 22:31:09 DONE info@ecuadorhost.com 0 11 view
10 test9.april15 04/15/2010 22:31:17 DONE info@ecuadorhost.com 0 11 view
num. rows: 10
>
>>
page number: 1
Done
Internet | Protected Mode: On
100%

Figure 8 Reports Web Page. Administration Interface

3.2.6 Analysis Request Detail

The analysis requests detail section shows the complete information about a particular request. The following figure shows the Web Page with the distribution of the information for one request.

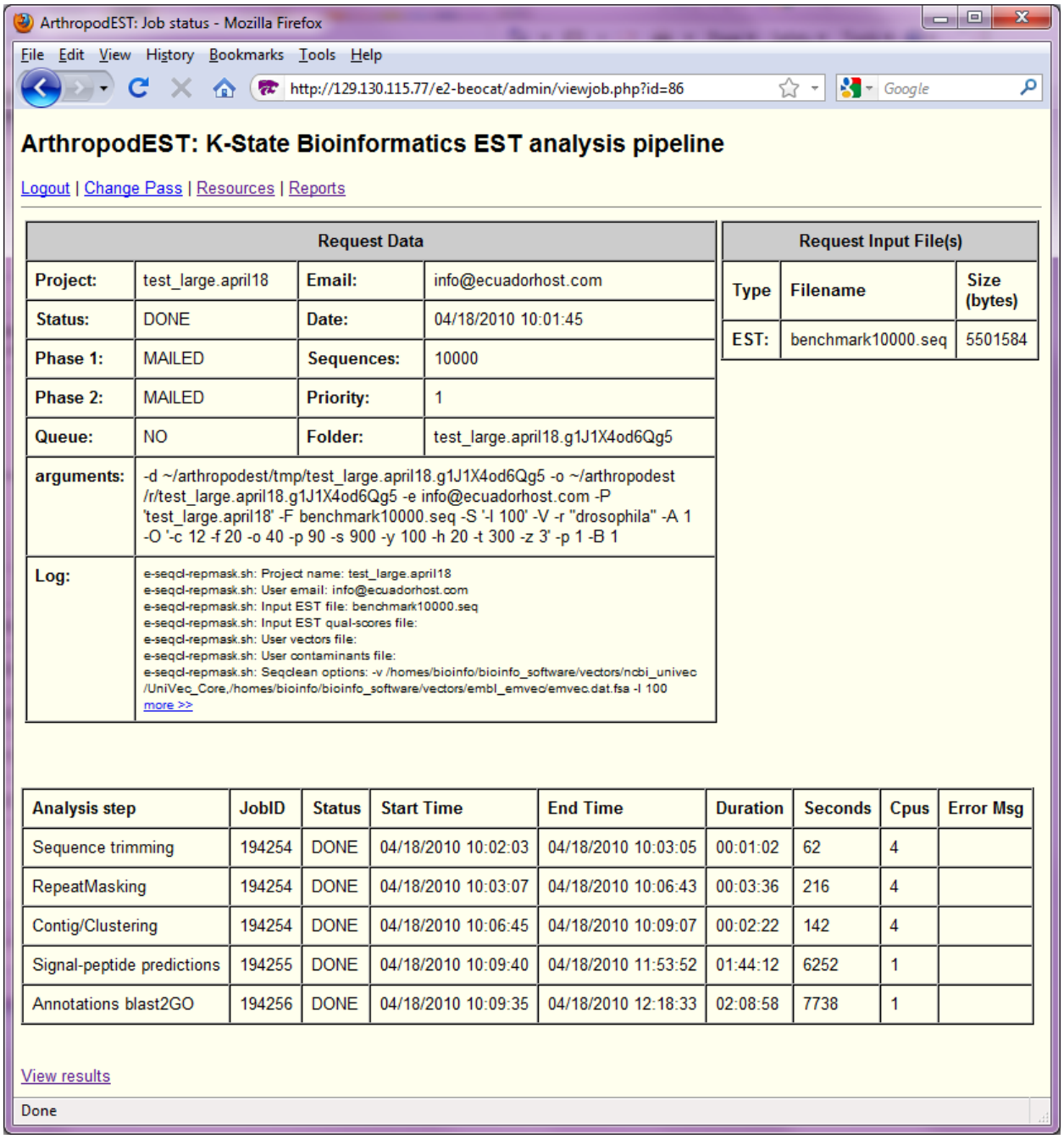
ArthropodEST: Job status - Mozilla Firefox
File Edit View History Bookmarks Tools Help
http://129.130.115.77/e2-beocat/admin/viewjob.php?id=86
Google
ArthropodEST: K-State Bioinformatics EST analysis pipeline
Logout | Change Pass | Resources | Reports
Request Data
Project: test_large.april18
Email: info@ecuadorhost.com
Status: DONE
Date: 04/18/2010 10:01:45
Phase 1: MAILED
Sequences: 10000
Phase 2: MAILED
Priority: 1
Queue: NO
Folder: test_large.april18.g1J1X4od6Qg5
arguments: -d ~/arthropodest/tmp/test_large.april18.g1J1X4od6Qg5 -o ~/arthropodest /r/test_large.april18.g1J1X4od6Qg5 -e info@ecuadorhost.com -P 'test_large.april18' -F benchmark10000.seq -S '-l 100' -V -r "drosophila" -A 1 -O '-c 12 -f 20 -o 40 -p 90 -s 900 -y 100 -h 20 -t 300 -z 3' -p 1 -B 1
Log:
e-seqcl-repmask.sh: Project name: test_large.april18
e-seqcl-repmask.sh: User email: info@ecuadorhost.com
e-seqcl-repmask.sh: Input EST file: benchmark10000.seq
e-seqcl-repmask.sh: Input EST qual-scores file:
e-seqcl-repmask.sh: User vectors file:
e-seqcl-repmask.sh: User contaminants file:
e-seqcl-repmask.sh: Seqclean options: -v /homes/bioinfo/bioinfo_software/vectors/ncbi_univec /UniVec_Core,/homes/bioinfo/bioinfo_software/vectors/embl_emvec/emvec.dat.fsa -l 100
more >>
Request Input File(s)
Type Filename Size (bytes)
EST: benchmark10000.seq 5501584
Analysis step JobID Status Start Time End Time Duration Seconds Cpus Error Msg
Sequence trimming 194254 DONE 04/18/2010 10:02:03 04/18/2010 10:03:05 00:01:02 62 4
RepeatMasking 194254 DONE 04/18/2010 10:03:07 04/18/2010 10:06:43 00:03:36 216 4
Contig/Clustering 194254 DONE 04/18/2010 10:06:45 04/18/2010 10:09:07 00:02:22 142 4
Signal-peptide predictions 194255 DONE 04/18/2010 10:09:40 04/18/2010 11:53:52 01:44:12 6252 1
Annotations blast2GO 194256 DONE 04/18/2010 10:09:35 04/18/2010 12:18:33 02:08:58 7738 1
View results
Done

Figure 9 Request Detail Web Page. Administration Interface

The section 'Request Data' consists of the main information regarding the analysis requests. The 'Request Input File(s)' section shows the input file(s) provided to perform the analysis. The table at the bottom shows its operations and the status of their execution. They include complete information about

every execution including start and end time, duration, number of CPUs used by Beocat and an error message, if any, produced while executing pipeline scripts.

3.2.7 Cancel Requests

During the execution of a request, the Administrator is able to cancel it using the Analysis Request Detail Web Page. As is shown Figure 10, a 'Cancel Request' button will be displayed and after pressing the button a message will appear so that Administrator confirms the intention to cancel the request. After cancelling, a notification will be sent to user indicating that the request was successfully cancelled

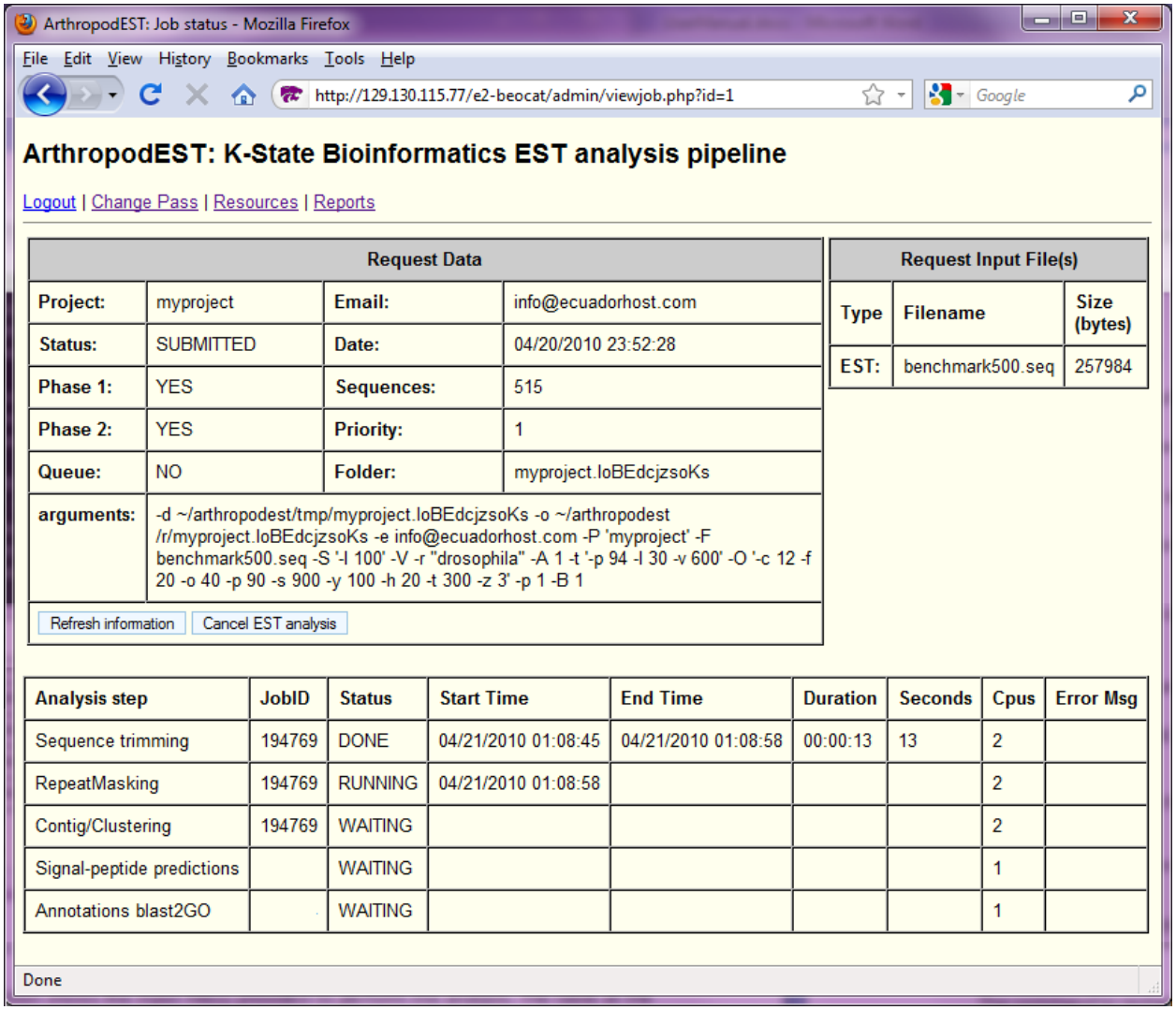
ArthropodEST: Job status - Mozilla Firefox
File Edit View History Bookmarks Tools Help
http://129.130.115.77/e2-beocat/admin/viewjob.php?id=1
Google
ArthropodEST: K-State Bioinformatics EST analysis pipeline
Logout | Change Pass | Resources | Reports
Request Data
Project: myproject
Email: info@ecuadorhost.com
Status: SUBMITTED
Date: 04/20/2010 23:52:28
Phase 1: YES
Sequences: 515
Phase 2: YES
Priority: 1
Queue: NO
Folder: myproject.loBEdcjzsoKs
arguments: -d ~/arthropodest/tmp/myproject.loBEdcjzsoKs -o ~/arthropodest /r/myproject.loBEdcjzsoKs -e info@ecuadorhost.com -P 'myproject' -F benchmark500.seq -S '-l 100' -V -r "drosophila" -A 1 -t '-p 94 -l 30 -v 600' -O '-c 12 -f 20 -o 40 -p 90 -s 900 -y 100 -h 20 -t 300 -z 3' -p 1 -B 1
Refresh information
Cancel EST analysis
Request Input File(s)
Type Filename Size (bytes)
EST: benchmark500.seq 257984
Analysis step JobID Status Start Time End Time Duration Seconds Cpus Error Msg
Sequence trimming 194769 DONE 04/21/2010 01:08:45 04/21/2010 01:08:58 00:00:13 13 2
RepeatMasking 194769 RUNNING 04/21/2010 01:08:58 2
Contig/Clustering 194769 WAITING 2
Signal-peptide predictions WAITING 1
Annotations blast2GO WAITING 1
Done

Figure 10 Cancel Request. Administration Interface

3.2.8 View Request Results

Also, on Analysis Request Detail Web Page, if results are ready to download, a link in the bottom of the page will allow users to open the Results Web Page. A screen shot is shown in Figure 4.

3.2.9 Start and Stop Engine coordinator

After the installation and configuration of ArthropodEST Pipeline system, the administrator must run apache web server on BC Server and MySQL before starting the engine component.

The command to start the engine component service is as follows:

```
python /path/to/engine/engine-est.py start
```

The command to stop the service is as follows:

```
python /path/to/engine/engine-est.py stop
```

The command to restart the service is as follows:

```
python /path/to/engine/engine-est.py restart
```